THE SEISMOLOGICAL SOCIETY OF AMERICA
400 Evelyn Ave., Suite 201
Albany, CA 94706-1375
(510) 525-5474; FAX (510) 525-7204
www.seismosoc.org

Bulletin of the Seismological Society of America, Vol. 104, No. 2, pp. 809–822, April 2014, doi: 10.1785/0120130105

Adaptive Smoothing of Seismicity in Time, Space, and Magnitude for Time-Dependent Earthquake Forecasts for California

by Agnès Helmstetter and Maximilian J. Werner

Abstract We present new methods for short-term earthquake forecasting that employ space, time, and magnitude kernels to smooth seismicity. These methods are purely statistical and rely on very few assumptions about seismicity. In particular, we do not use Omori–Utsu law, and only one of our two new models assumes a Gutenberg–Richter law to model the magnitude distribution; the second model estimates the magnitude distribution nonparametrically with kernels. We employ adaptive kernels of variable bandwidths to estimate seismicity in space, time, and magnitude bins. To project rates over short time scales into the future, we simply assume persistence, that is, a constant rate over short time windows. The resulting forecasts from the two new kernel models are compared with those of the epidemic-type aftershock sequence (ETAS) model generated by Werner *et al.* (2011). Although our new methods are simpler and require fewer parameters than ETAS, the obtained probability gains are surprisingly close. Nonetheless, ETAS performs significantly better in most comparisons, and the kernel model with a Gutenberg–Richter law attains larger gains than the kernel model that nonparametrically estimates the magnitude distribution. Finally, we show that combining ETAS and kernel model forecasts, by simply averaging the expected rate in each bin, can provide greater predictive skill than ETAS or the kernel models can achieve individually.

Introduction

Most short-term earthquake forecasts are based on empirical laws describing the statistical distribution of earthquakes in space, time, and magnitude. These empirical laws, such as the Omori–Utsu law for the temporal decay of aftershocks, are derived from statistical analyses of earthquake catalogs. These models usually assume that the seismicity rate is the sum of a background rate (usually heterogeneous in space and stationary in time) and of triggered earthquakes. This class of models includes, among others, the point-process model of Kagan and Knopoff (1987), the epidemic-type aftershock sequence (ETAS) model (Ogata, 1988; Helmstetter *et al.* 2006), the STEP model (Gerstenberger *et al.*, 2005), and the model of Marsan and Lengliné (2008).

Although the empirical laws appear to capture the first-order or average characteristics of earthquake clustering well, much evidence has been claimed to show that seismicity frequently deviates from this gross description (e.g., Vidale *et al.*, 2006; Ben-Zion, 2008; Enescu *et al.*, 2009; Shearer, 2012; Hainzl, 2013). To capture potential deviations, we wish to develop more flexible, purely data-driven models. For example, the Omori–Utsu law is widely observed to hold during aftershock sequences (e.g., Utsu *et al.*, 1995), but variations in the parameters (e.g., Kisslinger and Jones, 1991) and non-Omori-like behavior (e.g., Vidale *et al.*, 2006) have been noted. Similarly, whereas the Gutenberg–Richter law with exponent $b \approx 1$ approximates the magnitude distribution in sufficiently large space–time volumes well (e.g., Kagan, 1997), b-values have been claimed to vary at smaller scales (e.g., Wiemer *et al.*, 1998; Wiemer and Wyss, 2002). One approach to account for these seismicity fluctuations in models is to continuously recalibrate Omori's p-value and/or Gutenberg–Richter's b-value to the data to improve forecast performance (e.g., Gerstenberger *et al.*, 2005; Wiemer and Schorlemmer 2007). For instance, in the STEP model (Gerstenberger *et al.*, 2005), the parameters p and b are inverted for during aftershock sequences.

In this study, we use a different approach. Instead of building a complex model that accounts for variable model parameters, we develop simple, nonparametric models that are mostly based on kernel smoothing. We use the methods of Helmstetter and Werner (2012) to smooth seismicity in space and time using adaptive kernels. We test different versions of our models and compare their probability gains to those of the ETAS model of Werner *et al.* (2011). Our first model (K^2) uses adaptive smoothing in time and space to estimate the seismicity rate in each cell and a Gutenberg–Richter law with fixed b-value to model the magnitude distribution. In addition, we assume a stationary background

rate to better model earthquake density in zones of very sparse seismicity and to account for surprise earthquakes. This background rate was estimated by Helmstetter and Werner (2012) based on adaptive smoothing of instrumental seismicity. The second model (K^3), in addition to adaptively smoothing seismicity in time and space like model K^2, relaxes the assumption of a Gutenberg–Richter magnitude distribution and instead uses kernels to estimate the magnitude distribution by smoothing the magnitudes of nearby earthquakes. As for model K^2, the background rate is stationary, and its magnitude density obeys the Gutenberg–Richter law.

In this article, we first describe the different models used to estimate the seismicity rate in space, time, and magnitude, and their use for earthquake forecasting. We then calibrate all models to California seismicity, compare their predictive skills, and evaluate the influence of minimum magnitude thresholds and forecast horizons, that is, the time interval for which forecasts are prepared.

Earthquake Forecasts Based on Adaptive Kernels

Spatiotemporal Estimation of Seismicity Rate Using Adaptive Kernels (Model K^2)

Using adaptive kernels in time and space, the seismicity rate at time t and location $\mathbf{r}$ for events of magnitudes $\mathrm{M} \geq M_t$ can be written as a function of the times and locations of past earthquakes (Choi and Hall, 1999; Helmstetter and Werner, 2012):

$$R(\mathbf{r}, t) = \mu(\mathbf{r}) + f \sum_{t_i < t} \frac{2w_i}{h_i d_i^2} K_t\left(\frac{t - t_i}{h_i}\right) K_r\left(\frac{|\mathbf{r} - \mathbf{r_i}|}{d_i}\right), \quad (1)$$

in which h_i and d_i are the bandwidths associated with event i in the temporal and spatial domain, respectively, $|\mathbf{r} - \mathbf{r_i}|$ is the distance to the epicenter of event i, K_t is the temporal kernel and K_r is the spatial kernel. We use a univariate Gaussian kernel for K_t but modified it so that $K_t(t) = 0$ for $t < 0$, because $R(\mathbf{r}, t)$ cannot depend on future events for purposes of forecasting. In the spatial domain, we tested both a 2D Gaussian kernel and a power-law kernel. The factor 2 in equation (1) accounts for the fact that future events (with $t_i > t$) are missing in the sum. The stationary rate $\mu(\mathbf{r})$ assures that $R(\mathbf{r}, t)$ is always positive and also accounts for surprises, that is, earthquakes that occur in areas in which there were no events in the catalog. This rate $\mu(\mathbf{r})$ is chosen to be proportional to the long-term forecast estimated by Helmstetter and Werner (2012). Specifically, we define the stationary rate μ_0 by $\mu(\mathbf{r}) = \mu_0 \mu_r(\mathbf{r})$, in which $\mu_r(\mathbf{r})$ is the long-term rate normalized so that its integral over the testing area equals one. The weights w_i in equation (1), defined below, account for fluctuations of the completeness magnitude in time and space and for the fact that the minimum magnitude M_t of target events may be different from that of the learning catalog M_D. A corrective factor f in equation (1) is introduced to match the observed number of target events.

Following Helmstetter and Werner (2012), we use the coupled near-neighbor method proposed by Choi and Hall (1999) to estimate the kernel bandwidths in time h_i and space d_i. This method has two parameters, an integer k describing the overall level of smoothing and another parameter a controlling the relative importance of smoothing in space and time. For each earthquake i, the bandwidths d_i and h_i are estimated so as to minimize $h_i + a d_i$ under the constraint that there are k events at a distance smaller than d_i and at a time between $t_i - h_i$ and t_i. This method is illustrated in figure 1 of Helmstetter and Werner (2012). Coupling the temporal and spatial domains provides better results than estimating h_i and d_i independently from the distance and time between earthquake i and its closest neighbors in time and space. A larger value of a provides a forecast with higher resolution in space that is smoother in the time domain. Because of the limited accuracy of earthquake location, we impose that the bandwidth d_i cannot be smaller than 0.5 km. A different method can be used to estimate the bandwidths, namely from a pilot estimate of the density (Abramson, 1982; Helmstetter and Werner, 2012). For long-term forecasts, Helmstetter and Werner (2012) found that both methods yield similar results. However, for time-dependent forecasts we found that the near-neighbor method provides significantly better results, and we therefore do not describe the pilot-density method in this article.

In this first model K^2, we consider that the magnitude distribution obeys a tapered Gutenberg–Richter distribution. Following Helmstetter *et al.* (2007), we assume that the cumulative magnitude distribution has the form

$$P(\mathrm{M}) = 10^{-b(\mathrm{M}-M_t)} \exp[10^{1.5(M_t - M_\mathrm{c})} - 10^{1.5(\mathrm{M}-M_\mathrm{c})}], \quad (2)$$

with a corner magnitude $M_\mathrm{c} = 8$ and an exponent $b = 1$, as suggested by Bird and Kagan (2004) for continental transform fault boundaries. The exponential factor in equation (2) describes a fall-off of the distribution of seismic moments for $\mathrm{M} > M_\mathrm{c}$. In addition, following Helmstetter *et al.* (2007), Werner *et al.* (2011), and Helmstetter and Werner (2012), we introduce a correction for the region of the Geysers, a hydrogeothermal area in northern California. In this region, we assume a piecewise Gutenberg–Richter law with an exponent $b = 1$ for $\mathrm{M} < 3.4$ and $b = 2$ for $\mathrm{M} \geq 3.4$. This Gutenberg–Richter magnitude distribution is used to extrapolate the rate of $\mathrm{M} \geq M_t$ earthquakes from the rate of $\mathrm{M} \geq M_\mathrm{D}$ earthquakes in the learning catalog and to correct for variations of the completeness magnitude $M_0(t, \mathbf{r})$.

The weights w_i are estimated by

$$w_i = 10^{b[M_0(t_i, \mathbf{r_i}) - M_t]}, \quad (3)$$

in which $M_0(t_i, \mathbf{r_i})$ is the completeness magnitude estimated at the time and location of earthquake i. We use the spatial completeness magnitude map estimated by Werner *et al.* (2011) from the local magnitude distribution using adaptive kernels. In addition, M_0 increases following large $\mathrm{M} \geq 5$

earthquakes, and we follow the method proposed by Helmstetter *et al.* (2006) to model changes in the detection threshold after mainshocks with $M \geq 5$ such that

$$M_0(t) = M - 0.76 \log_{10}(t) - 4.5, \quad (4)$$

in which t is the time (in days) since an earthquake of magnitude M.

Estimation of Seismicity Rate in Time, Space, and Magnitude Using Adaptive Kernels (Model K^3)

In contrast to model K^2, we do not impose a magnitude distribution in model K^3. Instead, we use a magnitude kernel K_M to nonparametrically estimate the magnitude distribution, as in the time and space domains. We chose a Gaussian function for K_M for simplicity, as in the time domain. The expected rate of earthquakes at time t, location r, and with magnitude M is thus modeled by

$$R(\mathbf{r}, t, M) = \mu(\mathbf{r}, M) + f \sum_{t_i < t} \frac{2}{h_i d_i^2 \sigma_i} K_t\left(\frac{t - t_i}{h_i}\right) \times K_r\left(\frac{|\mathbf{r} - \mathbf{r_i}|}{d_i}\right) K_M\left(\frac{M - M_i}{\sigma_i}\right). \quad (5)$$

The stationary rate $\mu(\mathbf{r}, M)$ is again assumed to be proportional to the long-term model of Helmstetter and Werner (2012), as for our first model. Additionally, we assume for simplicity, and also to avoid empty magnitude bins, that $\mu(\mathbf{r}, M)$ decreases with M following a tapered Gutenberg–Richter law (equation 2). We tested two methods to estimate the bandwidth σ_i of the magnitude kernel. We used either a fixed bandwidth $\sigma_i = \sigma_0$ or a bandwidth that increases with magnitude according to

$$\sigma_i = \sigma_0 10^{(M_i - M_t)/2}. \quad (6)$$

This relation (6) is based on a result by Abramson (1982) that the kernel bandwidth should scale as the inverse of the square root of density. However, we found that a constant bandwidth provides better results (larger forecast likelihood) than using relation (6). In the following, we only present results obtained with a fixed magnitude bandwidth.

One general problem when using kernels to estimate probability density functions is the presence of boundary effects. In this study, we have no problems in the time or space domains because we can use available earthquakes outside of the forecast testing interval or region. However, in the magnitude dimension, side effects can produce an artificial roll-off for magnitudes close to the minimum magnitude M_D. This is problematic when the minimum target magnitude M_t is close to the minimum learning magnitude M_D. Several methods have been proposed to deal with this problem (Müller, 1991). In this work, we use the reflection technique introduced by Schuster (1985) to correct for missing earthquakes with $M \leq M_D$: for each earthquake of magnitude $M \geq M_D$ we add a mirror earthquake with magnitude equal to $M_D - (M - M_D) = 2M_D - M$. This method underestimates the rate of events close to M_D, because there are more events with $M < M_D$ than with $M > M_D$. Another way to deal with side effects for $M \approx M_D$ is to use a kernel bandwidth $\sigma(M)$ depending on magnitude, which decreases toward 0 for $M = M_D$ (Müller, 1991). This method works correctly even if the density function is not uniform close to the boundary. We tested both approaches and found that the reflection technique works best, although the differences in probability gain are very small. However, the main interest of earthquake forecasting is to use small earthquakes to predict larger ones ($M_t > M_D$). If $M_t - M_D \gg \sigma_0$, the estimated rate of $M > M_t$ earthquakes is weakly influenced by missing earthquakes with $M < M_D$ and the results do not depend on the method used to correct for side effects.

The choice of kernels to estimate the magnitude distribution is motivated by several arguments. First, nonparametric estimation can map spatial variations of the magnitude distribution, for example, without *ad hoc* corrections for the Geysers area. Second, the method adapts to spatial and temporal variations of the completeness magnitude. As a result, we do not calculate weights w_i to model spatiotemporal incompleteness. Finally, it can also account for potential correlations between earthquake magnitudes. For example, Lippiello *et al.* (2012) reported significant correlations between magnitudes for earthquakes close in time and space.

Short-Term Earthquake Forecasts Based on Adaptive Smoothing

To estimate the expected seismicity rate over a short forecast horizon, we simply assume that the future seismicity rate equals the current rate estimated by equation (1) or (5). That is, the kernel models provide simple persistence forecasts. Model K^2 has only four adjustable parameters: k, a, f, and μ_0. For K^3, we also invert for σ_0. For both models, the corrective factor f is estimated independently of the other parameters by constraining the total number of predicted events to match the number of target events.

Forecasts are issued at regular time intervals T. The predicted rates are integrated in time, space, and magnitude to compute the predicted number of events in each space, time, and magnitude bin.

Short-Term Earthquake Forecasts Based on the ETAS Model

We compare our new models with the ETAS model of Werner *et al.* (2011). In this model, the expected rate of earthquakes at time t, location r, and with magnitude equal to or greater than M can be written as a function of past earthquakes i as

$$R(\mathbf{r}, t, M) = P(M)[\mu(\mathbf{r}) + \sum_{t_i < t} K 10^{\alpha(M_i - M_t)} \phi(t - t_i) \psi(\mathbf{r} - \mathbf{r_i}, M_i)], \quad (7)$$

in which the temporal kernel $\phi(t)$ is the normalized Omori–Utsu law:

$$\phi(t) = \frac{(p-1)c^{p-1}}{(t+c)^p}. \quad (8)$$

For small triggering earthquakes M < 5.5, we model the spatial distribution of aftershocks by a power-law kernel

$$\psi(|\mathbf{r}|, \mathrm{M}) = \frac{C}{(|\mathbf{r}| + d_0 10^{\mathrm{M}/2})^{1.5}}, \quad (9)$$

in which C is a normalizing constant so that the integral over space of the spatial kernel equals one. The characteristic triggering distance $d_0 10^{\mathrm{M}/2}$ in equation (9) is proportional to the rupture length.

For larger earthquakes, we smooth the locations of early aftershocks (within three days) to map the spatial distribution of aftershocks and to estimate the function $\psi(\mathbf{r}, \mathrm{M})$ for each earthquake with M ≥ 5.5, following the method by Helmstetter *et al.* (2006).

We made minor changes to the ETAS model implemented by Werner *et al.* (2011). First, we no longer correct for changes in the completeness magnitude with time and space. This correction did not improve the performance of the forecasts but did increase computational time considerably. Second, we now invert for all parameters α, K, p, c, μ_0, and d_0, while Helmstetter *et al.* (2006) and Werner *et al.* (2011) fixed the cutoff time c of the Omori–Utsu law to 5 min.

We use the same magnitude distribution $P(\mathrm{M})$ (equation 2) as for model K^2, including the corrections in the Geysers area. The background rate $\mu(\mathbf{r}, \mathrm{M})$ is also proportional to the long-term model of Helmstetter and Werner (2012), as for the other models.

The definition (7) of the ETAS model is similar to equation (1) that defines kernel model K^2. A first difference is the shape of the temporal kernel: the ETAS model uses a power-law kernel whereas the kernel models use a Gaussian kernel. Another difference is that the ETAS model gives more importance to larger earthquakes, due to the term $10^{\alpha \mathrm{M}}$ in definition (7) that models the increase of aftershock productivity as a function of the mainshock magnitude. In contrast, the kernel models give the same weight to all earthquakes, irrespective of their magnitude. The weights w_i correct for catalog incompleteness and differences in target and learning magnitude thresholds only.

But the main difference between ETAS and kernel models is that ETAS is, in a sense, a physical model, whereas kernel models are purely statistical. That is, the kernels $\phi(t)$ and $\psi(\mathbf{r}, \mathrm{M})$ of the ETAS model are a statistical expression of presumed, underlying physical mechanisms of earthquake triggering. In contrast, the kernel models make no (or fewer) assumptions about the macroscopic expression of physical mechanisms responsible for the observed clustering between earthquakes; nonparametric estimation can be applied to any stochastic spatiotemporal point process.

Likelihood and Probability Gains

For all models, the parameters are optimized via maximum likelihood and assuming a time-dependent Poisson process. This implies that future events are assumed to be independent from each other. This assumption is clearly invalid if a large aftershock sequence happens within the time interval of the forecast. An alternative to the Poisson assumption for ETAS forecasts was proposed by Marzocchi and Lombardi (2009): the distribution of the number of events in each space–time bin is estimated from simulated earthquake catalogs. In the case of the kernel models, we do not have a method for simulating catalogs, and we therefore assume a Poisson process. The log likelihood of a Poisson process can be estimated by (e.g., Schorlemmer *et al.*, 2007)

$$L = \sum_{i_t=1}^{n_T} \Big\{ -N_p(i_t) + \sum_{i_x} \sum_{i_y} \sum_{i_M} [N(i_t, i_x, i_y, i_M) \times \log[N_p(i_t, i_x, i_y, i_M)] - \log(N(i_t, i_x, i_y, i_M)!)] \Big\}, \quad (10)$$

in which $N_p(i_t, i_x, i_y, i_M)$ is the number of predicted events in the cell (i_x, i_y, i_M) and time interval i_t (between $t(i_t)$ and $t(i_t) + T$); $N(i_t, i_x, i_y, i_M)$ is the observed number of earthquakes in this bin; $N_p(i_t)$ is the total number predicted for this time interval; and n_T is the number of time intervals. Note that the second term in the sum is zero for empty cells ($N(i_t, i_x, i_y, i_M) = 0$). We thus only need to compute this term for nonempty cells and can accelerate the computation.

We maximize the likelihood using a simplex algorithm and compare it with the likelihood of a reference model. Our reference model assumes a uniform Poisson process (in space and time) and a tapered Gutenberg–Richter law (equation 2) with the same corner magnitude and b-value as K^2 and ETAS, but without the special correction in the Geysers area. The reference model assumes a constant rate per cell, therefore the density (predicted number of events in each cell divided by the cell area) is not exactly uniform.

The probability gain G is used to compare our models with the reference model

$$G = \exp\left(\frac{L - L_r}{N}\right), \quad (11)$$

in which L_r is the log likelihood of the reference model and N is the total number of target events.

When comparing two models that expect the same total number of events, the prediction gain represents the geometric average of the ratio of the predicted rate for each model (Helmstetter and Werner, 2012, their equation 12)

$$G = \exp\left[\sum_{j=1,N} \frac{1}{N} \log\left(\frac{N_{p,1}[i_t(j), i_x(j), i_y(j), i_M(j)]}{N_{p,2}[i_t(j), i_x(j), i_y(j), i_M(j)]}\right)\right], \quad (12)$$

in which $N_{p,1}$ denotes the rate predicted by the first model in the space–time–magnitude bin $[i_t(j), i_x(j), i_y(j), i_M(j)]$ in which event j occurred.

Table 1
Parameters of ETAS and Kernel Models

			ETAS*						$K^{2\dagger}$				$K^{3\ddagger}$				
M_D	M_t	T	α	p	K	μ_0	d_0	c	k	a	f	μ_0	k	a	σ_0	f	μ_0
2	2	1	0.44	1.17	1.02	4.11	0.073	0.006	2	328	0.99	1.80	2	308	0.29	0.87	2.52
2	3	1	0.73	1.30	0.39	0.55	0.101	0.040	2	184	0.61	0.59	2	174	0.57	0.34	1.05
2	4	1	0.78	1.28	0.40	0.06	0.290	0.015	1	78	0.49	0.13	2	7	0.86	0.11	0.51
2	5	1	0.80	1.29	0.18	0.01	0.313	0.127	1	1174	0.42	0.02	1	334	1.27	0.02	0.41
3	3	1	0.62	1.20	0.47	0.67	0.147	0.024	2	83	0.69	0.85	2	85	0.37	0.67	0.87
3	4	1	0.70	1.23	0.40	0.07	0.263	0.019	2	83	0.63	0.10	2	33	0.71	0.24	0.27
3	5	1	0.76	1.20	0.20	0.01	0.006	0.059	1	1993	0.50	0.02	1	291	0.93	0.08	0.10
4	4	1	0.71	1.21	0.27	0.09	0.309	0.045	2	145	0.64	0.11	2	39	0.69	0.49	0.18
2	4	0.04	0.79	1.32	0.20	0.04	0.381	0.056	1	75	0.46	0.15	2	3.4	0.35	0.15	0.42
3	4	0.04	0.77	1.15	0.25	0.05	0.112	0.004	1	443	0.76	0.05	1	15	0.47	0.36	0.17
2	4	10	0.70	1.38	1.06	0.08	0.485	0.028	2	2147	0.52	0.13	6	70	0.82	0.11	0.77
3	4	10	0.56	1.35	1.41	0.09	0.428	0.017	2	12387	0.40	0.23	2	245	0.75	0.13	0.74
2	4	91	0.44	1.23	0.75	0.13	0.793	0.121	1	112379	0.31	0.39	8	159	0.78	0.08	1.43
3	4	91	0.55	1.41	0.49	0.13	0.603	1.954	2	21703	0.40	0.31	2	228	0.64	0.15	0.86

*ETAS parameters α, p, K, μ_0, and d_0 are defined in equations (7)–(9).
†Parameters k, a, f, and μ_0 are defined in the Spatiotemporal Estimation of Seismicity Rate Using Adaptive Kernels (Model K^2) section.
‡σ_0 is the bandwidth of the magnitude kernel.

In addition, we introduce the prediction gains G_t and G_M to evaluate the performance of models in time and magnitude. The gain G_t compares predicted rates in the time domain only, by integrating over magnitude and space. It is obtained by integrating the predicted rates $N_p[i_t(j), i_x(j), i_y(j), i_M(j)]$ in all grid cells and magnitude bins

$$G_t = \exp\left[\sum_{j=1,N} \frac{1}{N} \log\left(\frac{N_{p,1}[i_t(j)]}{N_{p,2}[i_t(j)]}\right)\right]. \quad (13)$$

Similarly, the probability gain G_M considers only the magnitude domain. It compares the magnitude distribution p_M for each target earthquake j with the reference magnitude distribution $p_{M,r}$

$$G_M = \exp\left[\sum_{j=1,N} \frac{1}{N} \log\left(\frac{p_M[i_M(j), i_t(j), i_x(j), i_y(j)]}{p_{M,r}[i_M(j)]}\right)\right], \quad (14)$$

in which

$$p_M(i_M, i_x, i_y, i_t) = \frac{N_p(i_M, i_x, i_y, i_t)}{\sum_{i=1,N_M} N_p(i, i_x, i_y, i_t)} \quad (15)$$

is the conditional magnitude distribution.

Application to California

We calibrated all models to Californian seismicity inside the region used by the Collaboratory for the Study of Earthquake Predictability (CSEP) (Schorlemmer and Gerstenberger, 2007) for California with a grid spacing of 0.1°. We tested different values of the minimum magnitude of earthquakes in the learning and target catalogs, as well as different forecast horizons T. We used the earthquake catalog compiled by the Advanced National Seismic System (ANSS) in the period from 1 January 1981 until 1 March 2012 with magnitude M ≥ 2 and depth smaller than 30 km (in accordance with CSEP's data collection). This is a composite catalog that gathers data from different seismic networks in the United States. We removed underground nuclear explosions at the Nevada Test Site following Werner *et al.* (2011). There are also problematic events which appear twice (or more often) in the catalog; these are mostly small events. We did not attempt to remove all the duplicate events; we only removed 71 events that had exactly the same time and almost identical location and magnitude as the preceding event. This leaves 190,058 events with latitudes ranging between 30.5° N and 44° N and longitudes between 112.1° W and 126.4° W. The catalog is reasonably complete above M 2 in the central part of California, but the completeness magnitude increases close to the boundaries. In addition, there are also transient increases of the completeness magnitude following large earthquakes. Following Werner *et al.* (2011), all earthquakes between 1 January 1986 and 1 March 2012 above the magnitude threshold M_t in the CSEP testing area are used as target earthquakes, and all events with M ≥ M_D since 1 January 1981 are used to compute the predicted number of events. Forecasts are issued at regular time intervals T, with T ranging between one hour and three months. All models estimate the rate of M ≥ M_t target events in each time, magnitude, and space intervals, following the rules defined by CSEP.

Results

We compared earthquake forecasts obtained with ETAS model and with kernel models K^2 (which assumes adaptive kernels in time and space and a Gutenberg–Richter law for magnitudes) and K^3 (kernels in time, space, and magnitude). The model parameters are given in Table 1 and the results are summarized in Table 2.

Table 2
Probability Gains for ETAS and Kernel Models

			ETAS			K^2			K^3		
M_D	M_t	T (days)	G	G_M	G_t	G	G_M	G_t	G	G_M	G_t
2	2	1	**47.40**	1.000	1.44	46.74	1.000	1.45	46.79	0.981	1.47
2	3	1	**52.12**	1.001	1.65	47.57	1.001	1.62	47.35	0.965	1.63
2	4	1	**47.37**	1.002	1.67	41.93	1.002	1.65	41.32	0.981	1.64
2	5	1	**23.37**	0.999	1.34	22.14	1.000	1.36	22.08	0.984	1.35
3	3	1	**45.18**	1.001	1.62	40.57	1.001	1.60	38.32	0.948	1.64
3	4	1	**42.76**	1.002	1.66	34.61	1.002	1.62	34.64	0.977	1.63
3	5	1	**22.37**	0.999	1.33	18.72	1.000	1.23	20.68	0.988	1.26
4	4	1	**33.55**	1.002	1.65	22.82	1.002	1.57	20.83	0.929	1.54
2	4	0.04	**213.0**	1.002	2.56	133.1	1.002	2.31	94.04	0.798	2.23
3	4	0.04	**198.5**	1.002	3.23	119.1	1.002	2.69	98.22	0.851	2.67
2	4	10	**16.69**	1.002	1.20	13.87	1.002	1.16	14.64	0.988	1.20
3	4	10	**15.97**	1.002	1.19	12.84	1.002	1.11	13.67	0.987	1.12
2	4	91	6.483	1.002	1.01	6.521	1.002	0.99	**6.581**	0.999	1.01
3	4	91	**6.823**	1.002	1.02	6.301	1.002	0.99	6.387	0.996	1.00

Probability gains G, G_M, and G_t (see definition in Likelihood and Probability Gains section) for earthquake forecasts with different values of the minimum magnitude of target events M_t, of the minimum magnitude M_D of the learning catalog, and of the time interval T, computed for ETAS model and for kernel models K^2 and K^3. Bold values indicate the model with the largest gain G for each set of parameters.

Time Domain

Figure 1 compares the observed and predicted number of earthquakes per day following the two largest earthquakes in the catalog, the Landers M 7.3 28 June 1992 mainshock and the El Mayor–Cucapah M 7.2 10 April 2010 mainshock. All three models (ETAS, K^2, and K^3) shown in this figure were optimized to forecast M ≥ 3 targets using M ≥ 2 earthquakes. All models explain the decrease of aftershocks with

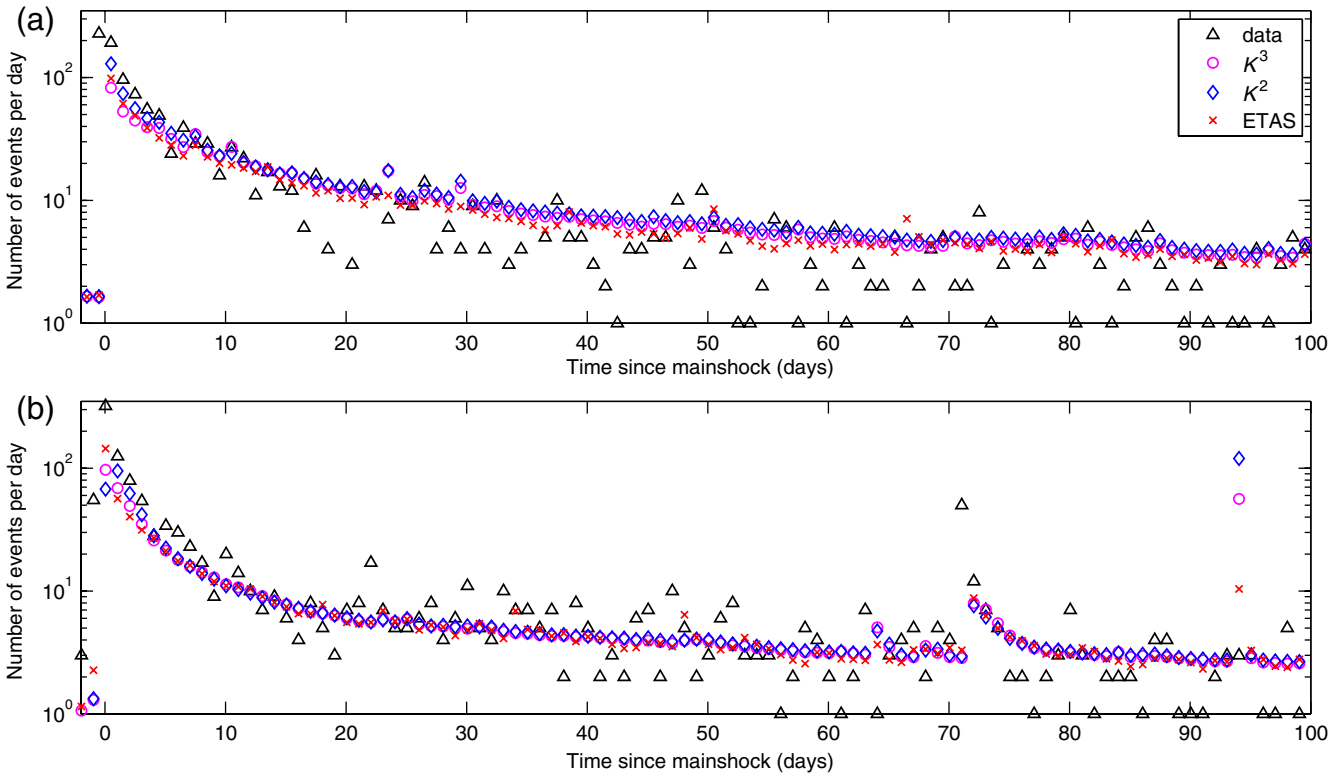

Figure 1. Observed (triangles) and predicted number of M ≥ 3 earthquakes per day following (a) the Landers M 7.3 28 June 1992 mainshock and (b) the El Mayor–Cucapah M 7.2 10 April 2010 mainshock. The predicted number of earthquakes per day is shown for ETAS (crosses), K^2 (diamonds), and K^3 (circles) models. The time axis begins on the day before each mainshock. The color version of this figure is available only in the electronic edition.

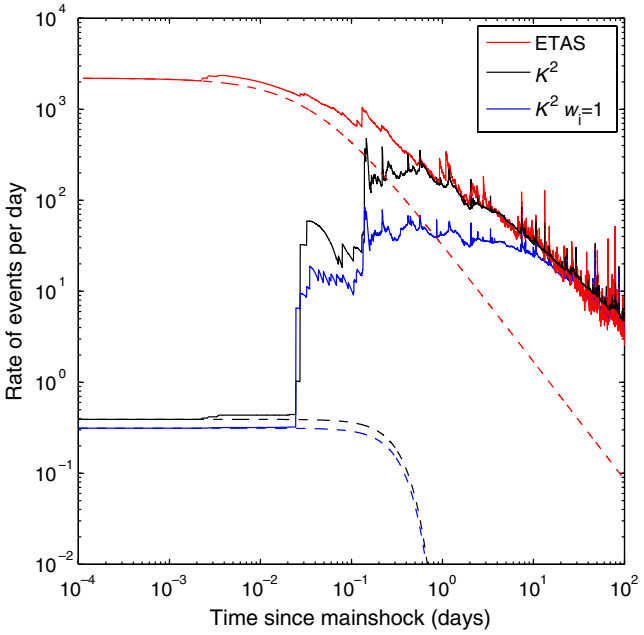

Figure 2. The predicted rate of M ≥ 3 earthquakes per day following the Landers M 7.3 28 June 1992 mainshock, for the ETAS and K^2 models using M ≥ 2 earthquakes. In addition, we also plotted the rate predicted by K^2 model by setting the same weight w_i = 1 for all events. This highlights the importance of correcting for missing early aftershocks using equation (3). The dashed lines represent the contribution of the mainshock, whereas the solid lines show the predicted rate estimated using all earthquakes that occurred after the mainshock. The color version of this figure is available only in the electronic edition.

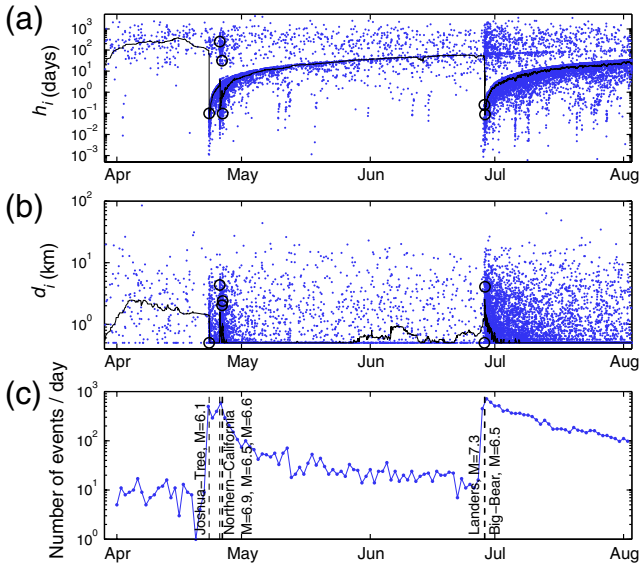

Figure 3. Temporal evolution of (a) temporal bandwidths, (b) spatial bandwidths, and (c) number of M ≥ 2 earthquakes per day for a period of four months in 1992, using the kernel model K^2 optimized for $T = 1$ day and $M_t = 3$. Circles in (a,b) indicate the occurrence of M ≥ 6 earthquakes. Continuous lines in (a,b) show the median value of kernel bandwidth in time and space computed in a sliding window of 100 events. The color version of this figure is available only in the electronic edition.

time reasonably well, but they all slightly underestimate the rate of aftershocks during the first five days after the mainshocks. In the case of Landers, this is especially true for K^3 and ETAS. The observed underprediction in these two sequences is not, however, emblematic of all sequences: there are mainshocks for which the predicted aftershock rate is larger than the observed rate. The predicted rate for the days of the Landers and El Mayor–Cucapah mainshocks is low for all models. The foreshock activity reported by Dodge *et al.* (1996) and Hauksson *et al.* (2010) did not increase the rate significantly by midnight of the day preceding these mainshocks (i.e., by the time forecasts are issued). This illustrates that larger probability gains can be achieved by reducing the forecast horizon to periods smaller than one day.

In Figure 1, there is a large peak of the predicted rate for all models observed on 8 July 2010, 94 days after the El Mayor–Cucapah mainshock. This peak of the predicted rate is not associated with a peak of the observed number of M ≥ 3 earthquakes. The peak is due to the occurrence of an M 5.4 earthquake on 7 July 2010 at 23:53 UTC. This earthquake was followed by 9 $m \geq 2$ mainshocks on 7 July 2010 within 7 min, yielding an average rate of 2236 events per day between the mainshock at 23:53 and midnight. This large rate of early aftershocks yields a large number of predicted $m \geq 3$ events using the kernel models, because these models assume that the rate for the day 8 July 2010 is equal to the rate measured at 00:00 on this day. Updating the models more often would remove much of this bias. For the ETAS model, the predicted rate for this day is much smaller than for the kernel models, because the ETAS model gives less weight to small earthquakes than kernel models.

Although the kernel models use a Gaussian temporal kernel, the predicted rates can well approximate a power-law decay of aftershocks with time. This is illustrated in Figure 2, which compares the temporal kernels of ETAS and K^2 models. In the first hours following the mainshock, model K^2 seriously underestimates the rate of aftershocks, because the weight associated with each earthquake does not depend on its magnitude and because the bandwidths of the temporal kernels are rather large. Figure 2 also shows the predicted rate of an unweighted version of model K^2, obtained by setting all weights $w_i = 1$ for all earthquakes in model K^2. At times shorter than 1 day, the predicted rate is further decreased by a factor of about 10 compared to the weighted K^2 model, demonstrating that the unweighted model does not account for the triggering due to missing early aftershocks. After several days, when the completeness magnitude decreases back to $M_0 = 2$, the unweighted rate gradually recovers the rate predicted by the weighted K^2 model. This highlights the importance of correcting for short-term fluctuations of the completeness magnitude using equation (3). Indeed, using weights to correct for catalog incompleteness can considerably increase the prediction gain, in particular when the magnitude of targets M_t is much larger than M_D. At later times, the

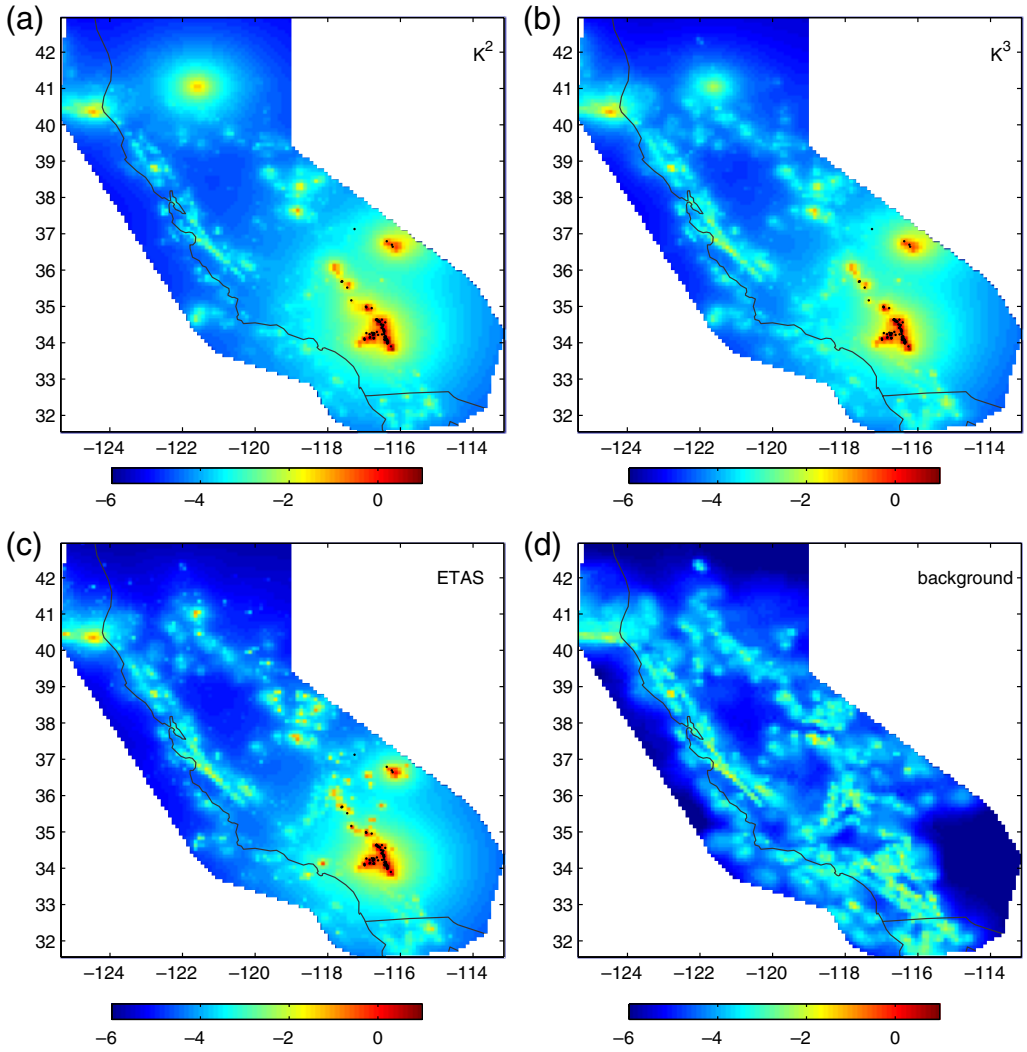

Figure 4. Locations of the predicted rate of M ≥ 3 earthquakes for 30 June 1992, two days after the Landers M 7.3 mainshock, for (a) K^2, (b) K^3, and (c) ETAS models, in logarithmic scale. Black dots represent M ≥ 3 target earthquakes that occurred on 30 June 1992. The background rate used by the ETAS model is shown in (d). Kernel models use the same background rate as ETAS but with a slightly different amplitude ($\mu_0 = 0.55$ background events per day for ETAS, 0.35 for K^2, and 0.39 for K^3). The color version of this figure is available only in the electronic edition.

temporal bandwidths have decreased a lot (from several days to less than 1 hour) and the high rate of activity allows for an accurate estimation of the seismicity rate. Figure 3 illustrates how the kernel bandwidths in space and time evolve with time and adapt to the seismicity rate. Just after a large earthquake, h_i can decrease to a few minutes and d_i to 0.5 km (the imposed lower boundary). In areas of low seismicity rate, in contrast, h_i can reach several years and d_i 100 km.

Kernel models provide a direct measure of the seismicity rate in space and time. Because the number of aftershocks per day decreases slowly with time (except on the first day), kernel models explain rather well the temporal evolution of aftershocks with time. This works because the learning catalog is updated every day, even if the Gaussian temporal kernel is very different from a power-law decay. In contrast, the ETAS model uses the Omori–Utsu law to model triggered earthquakes. It can thus predict the shape of the temporal decay of aftershocks correctly, even without updating the learning catalog.

Spatial Domain

For all models, we compared the Gaussian spatial kernel with a power-law kernel, and found that the power-law kernel gives generally better results. The difference in probability gain is small, however, generally of the order of 1%. The maps of predicted rate for the day 30 June 1992 (two days after the Landers mainshock) are shown in Figure 4. All models use the same background model but with different weights. For instance, for $M_D = 2$ and $M_t = 3$, the stationary rate is 0.55 earthquakes per day for ETAS model and for the whole testing area. For kernels models, the stationary rate is slightly smaller, with respectively 0.35 and 0.39 events per day for models K^2 and K^3. The maps for all three models are

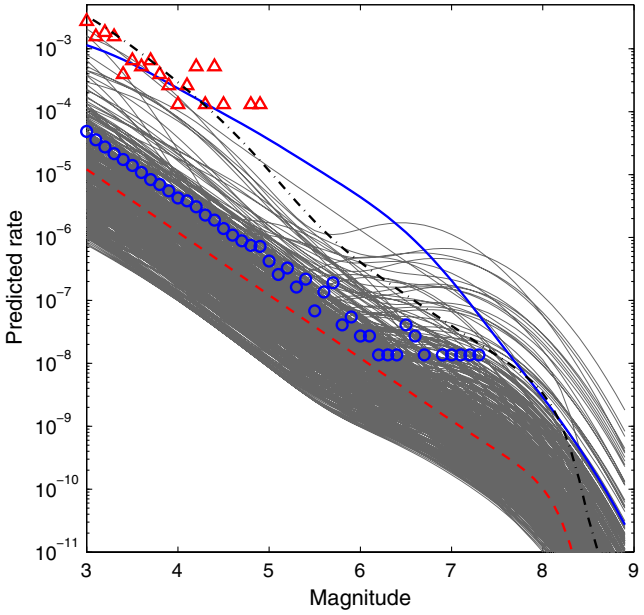

Figure 5. Observed and predicted number of earthquakes per magnitude bin, per day, and per cell, for model K^3 with $T = 1$ day, $M_D = 2$, and $M_t = 3$. The dashed line represents the contribution of the stationary rate, modeled by a tapered Gutenberg–Richter law with $b = 1$ and a corner magnitude $M_c = 8$. The predicted rate (averaged over all cells) for 30 June 1992 (two days after Landers mainshock) is shown by a solid line. The thin gray lines show the predicted rate for this day for a random selection of cells. The dashed-dotted line corresponds to the Geysers area. The circles represent the average rate of events per cell for the whole catalog, and the triangles show the average rate for the 96 earthquakes of 30 June 1992 (see text). The color version of this figure is available only in the electronic edition.

rather similar. All models predict a large increase in seismicity rate in the aftershock zone. The main difference is that kernel models are a little smoother: the kernel bandwidths in space and time are larger than the interevent times and distances. In contrast, ETAS model can predict more abrupt changes of seismicity rate.

Magnitude Distribution

Models ETAS and K^2 assume that the magnitude distribution obeys a tapered Gutenberg–Richter distribution, with a uniform exponent $b = 1$ (except in the Geysers area) and a corner magnitude $M_c = 8$. In contrast, model K^3 uses kernel smoothing of past earthquakes to estimate the magnitude distribution in each cell and at each time step. Only a small fraction of the seismicity is modeled by a stationary rate with a tapered Gutenberg–Richter distribution, especially during periods of active triggering. The magnitude distributions obtained by K^3 on day 30 June 1992 (two days after Landers mainshock) are shown in Figure 5. When averaging over all cells, the magnitude distribution is close to a Gutenberg–Richter law with $b \approx 1$ for small magnitudes but with a faster decay for $M \geq 6$. However, the magnitude distribution varies considerably between individual spatial cells. For example due to the occurrence of the M 7.3 Landers mainshock two days earlier, many cells show a relative increase near $M \approx 7$ over a Gutenberg–Richter distribution. Figure 5 also displays the magnitude distribution in the hydrothermal Geysers area. As expected, the magnitude distribution in this region (dashed-dotted black line) is steeper than the reference curve (dashed line), at least for magnitudes $M < 6$.

Comparison of the Predicted Number of Events per Day

Figure 6 compares the daily rates predicted by the ETAS and K^2 models on days when at least one target earthquake occurs. There is a good correlation between the two rates, but also significant differences. For example, on 8 July 2010, K^2 predicts a much larger rate than ETAS. This event was discussed previously in the Time Domain subsection. It is highlighted by a cross in Figure 6a and corresponds to day 94 in Figure 1b. The opposite behavior (ETAS predicting a much higher rate than K^2) is more frequent. An extreme example is shown by a triangle in Figure 6a. It corresponds to 17 March 2002, one day after an M 4.6 earthquake, in a cell located at 33.67° N and 119.33° W in which the seismicity rate is moderate. The M 4.6 earthquake occurred at 21:33, and only one $M \geq 2$ aftershock occurred on the same day. The rate predicted by K^2 for the following day was relatively small (7×10^{-5}). The bandwidths in space and time associated with these two earthquakes of 16 March 2002 are relatively large, thus not producing a significant increase of the predicted rate. In contrast, ETAS predicted a strong increase in seismicity due to the occurrence of the M 4.6 earthquake, because the rate depends exponentially on the magnitude. Most days on which K^2 underpredicts the seismicity rate can be explained similarly. Model K^2 needs at least k events in a small time and space interval to produce a strong increase in seismicity rate.

In these examples, ETAS performs better because the predicted rate strongly depends on the earthquake magnitude, unlike in the kernel models. However, on most days with at least one earthquake, kernel models are better than ETAS. Indeed, for $M_D = 2$, $M_t = 3$, and $T = 1$ day, the rate N_{p,K^2} predicted by the K^2 model in cells where targets occurred is larger than the rate predicted by the ETAS model for 56% of target earthquakes. The distribution of the ratio $x = N_{p,K^2}/N_{p,\mathrm{ETAS}}$ is shown in Figure 6b (for all nonempty bins). The arithmetic average and the median values of x are both slightly larger than 1, suggesting that model K^2 outperforms ETAS. However, the geometric average of x (equal to the probability gain defined in equation 11) is less than 1, indicating that, in terms of likelihood, the ETAS model is better.

To estimate the significance of the differences in probability gains, we could use the T-test or the W-test (Rhoades *et al.*, 2011). However, the empirical distribution of x points to issues in the interpretation of the results. For example, we found that the probability gain of the ETAS model is significantly larger than that of K^2 according to the T-test, whereas,

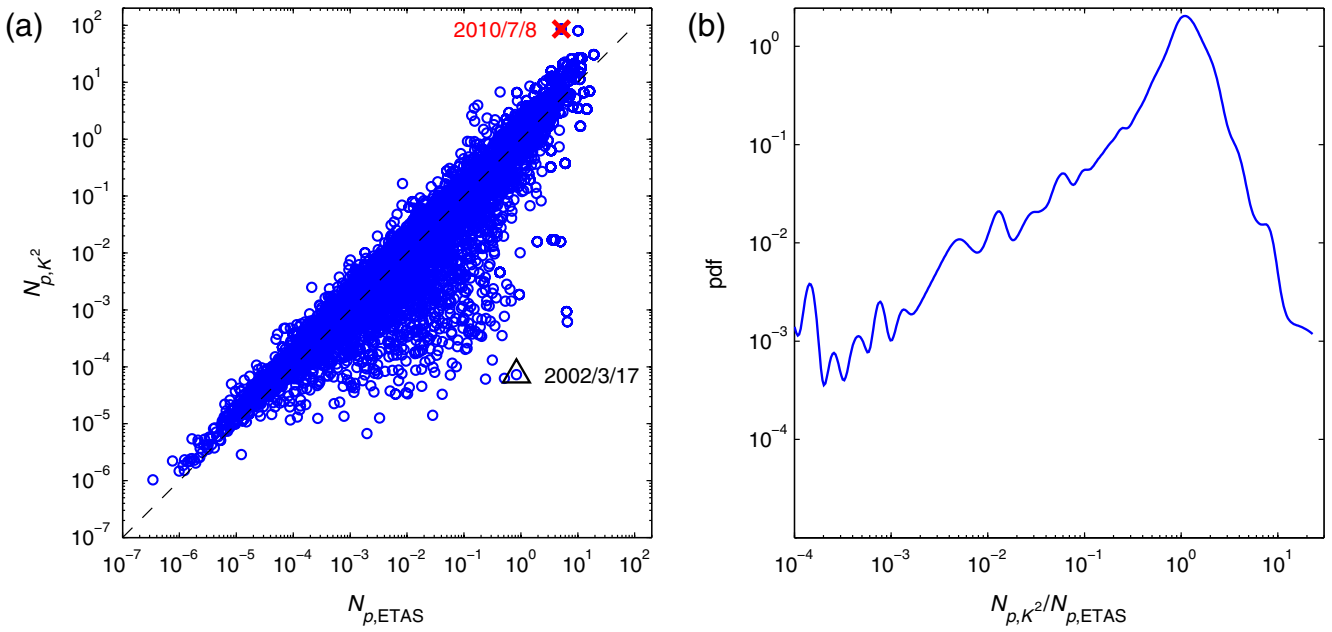

Figure 6. (a) Comparison of predicted rate per bin for models ETAS and K^2, using parameters $T = 1$ day, $M_\mathrm{D} = 2$, and $M_t = 3$. Only bins with at least one target event are plotted. (b) Probability density function of the ratio of the predicted rate for ETAS and K^2 models (thin line). The cross in (a) corresponds to earthquakes that occurred on 8 July 2010, and the triangle to earthquakes that occurred on 17 March 2002. The color version of this figure is available only in the electronic edition.

simultaneously, the W-test suggests that the median of the distribution significantly favors K^2. These two test results can both be true because the tests measure a particular property of a skewed and asymmetric empirical distribution. Deciding which model is better is thus a difficult task and depends on which quantity is considered: the median rate in the case of the W-test or the geometric average of the predicted rates (the probability gain) in the case of the T-test (see also the discussion by Eberhard *et al.*, 2012, in their section 7.1, and by Werner, Ide, and Sornette, 2011, in their section 4.2.2 and their fig. 6, in particular).

Kernel models generally perform better than ETAS when the seismicity rate varies smoothly, that is, at time and spatial scales larger than the interevent times and distances. In such cases, kernel models allow a direct and accurate measure of the seismicity rate in space and time. These models are not sensitive to the magnitudes of earthquakes, which are not always very accurate. In addition, there may be variations of the aftershock productivity or the Omori–Utsu law decay between sequences. When using the ETAS model, these variations can induce deviations between the observed and predicted seismicity rates. In contrast, kernel models are not sensitive to these fluctuations and can adapt to each situation, as long as there is no abrupt change of the seismicity rate in time or space.

Probability Gains as a Function of Minimum Magnitude and Forecast Horizon

We tested different values of the minimum magnitude M_D (for the learning catalog) and M_t (for targets), as well as different forecast horizons T. The results are summarized in Figure 7 and Table 2. In most cases, the ETAS model obtains larger probability gains G than kernel models. We find that the gain decreases with the minimum magnitude M_D, suggesting that including smaller events improves the forecasts. There is no clear trend of G as a function of the minimum magnitude of targets. The gain is roughly constant for $2 \le M_t \le 4$ but decreases for M_t 5, possibly due to the limited number of $M_t \ge 5$ targets. Updating the forecast less often (increasing T) strongly impacts the predictability. The gain decreases very fast with increasing T. For a time interval of 3 months, G is almost the same as the value $G_{LT} = 4.6$ obtained for the long-term (five-year) forecast by Helmstetter and Werner (2012).

The fact that ETAS model performs better than kernel models could be due partly to the larger number of inverted parameters for ETAS model. Indeed, the number of inverted parameters is six for ETAS model, five for K^3 model, and four for K^2 model. One way to account for such variations is the Akaike information criterion (AIC; Akaike, 1974), defined as $\mathrm{AIC} = 2(n - L) = 2(n - N\log(G)) + \text{constant}$, in which L is the log likelihood, n is the number of inverted parameters, N the number of targets, and G the probability gain. The preferred model is the one with the minimum AIC value. We found that comparing AIC values rather than gains do not change the results; ETAS is still the preferred model with the lowest AIC value for all forecast parameters tested in Table 2, except for the test with $M_\mathrm{D} = 2$, $M_t = 3$, and $T = 91$ days.

The probability gain in the time domain G_t is small compared to both G and the values of G_{LT} of long-term spatial

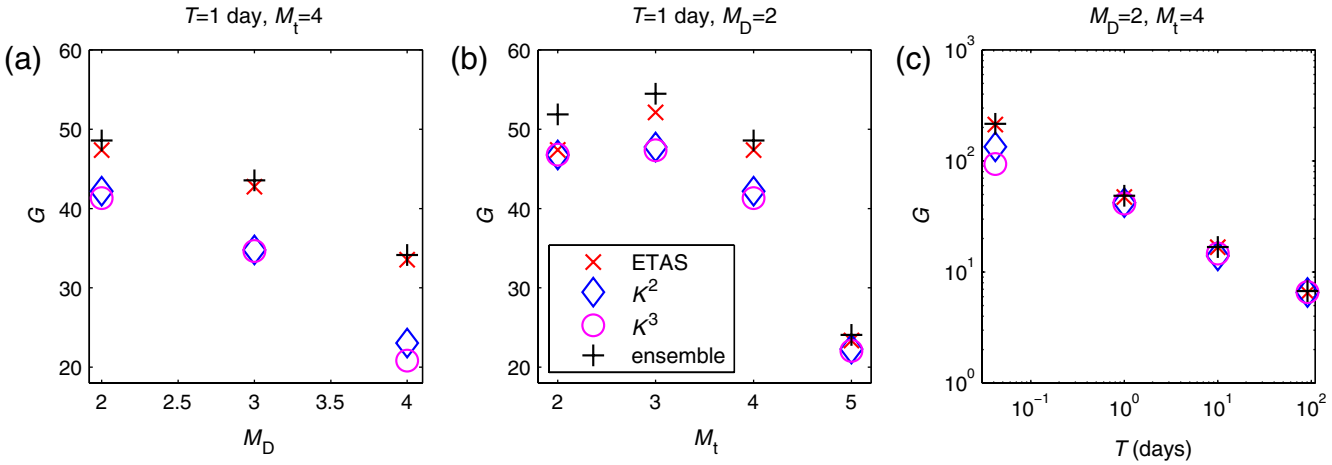

Figure 7. Probability gain as a function of (a) minimum magnitude M_{D} of the learning catalog, (b) minimum magnitude M_t of targets, and (c) time interval T for ETAS (crosses), K^2 (diamonds), and K^3 (circles) models. The best ensemble model (bold in Table 3) is shown by a plus. The color version of this figure is available only in the electronic edition.

models. Indeed, for $T \geq 1$ day, $G_t < 1.7$ (Table 2), demonstrating that the predictive skills of the models arise mostly from spatial information. Interestingly, modeling the spatial distribution of clustering in the ETAS model is difficult due to the strong anisotropy of aftershock distribution.

The probability gain G_M in the magnitude dimension is almost equal to one for models ETAS and K^2 for all cases shown in Table 2. The small deviation from one is due to the correction introduced in the Geysers area. For model K^3, the gain is always a little smaller than 1, suggesting that the magnitude distribution is better explained by a simple tapered Gutenberg–Richter law. The gain G_M increases with T because the optimized value of k (number of nearest neighbors) increases with T (Table 1). For $M_{\mathrm{D}} = 2$ and $M_t = 4$, the best values are $k = 2$ for $T = 1$ hr or 1 day, $k = 6$ for $T = 10$ days, and $k = 8$ for $T = 3$ months. As a consequence, the kernel bandwidths (h_i and d_i) also increase with T, so that there are more earthquakes with a significant contribution in equation (5). The gain G_M is also larger when $M_t > M_{\mathrm{D}}$, because boundary effects near M_{D} are less problematic.

Simple Ensemble Models

We constructed simple ensemble models of the different models by simply estimating a weighted average of the forecasted rate in each bin provided by two models. The predicted rate for an ensemble model is defined by

$$N_{p,\langle 1,2\rangle} = \omega N_{p,1} + (1-\omega) N_{p,2}, \tag{16}$$

in which $0 \leq \omega \leq 1$ is the mixing weight and $N_{p,1}$ and $N_{p,2}$ are the rates in each space–time–magnitude bin predicted by the two individual models. These average models generally perform better than each individual model, as shown in Table 3 and in Figure 7. The individual model with the largest gain has generally the largest weight in the ensemble model ($\omega > 0.5$), although there is an exception for the first line in Table 3, for $M_{\mathrm{D}} = 2$, $M_t = 2$, and $T = 1$ day.

In order to test whether ensemble forecasts are better despite having a larger number of parameters, we used the AIC test (Akaike, 1974). We consider that the number of free parameters of ensemble models are given by the sum of parameters of each individual model plus one for the optimal mixing weight ω. For each ensemble model, we compute the difference of AIC values δAIC between the ensemble model and the individual model with the smallest AIC value. These values are reported in Table 3. We found that in most cases the ensemble model is significantly better than each individual model ($\delta\mathrm{AIC} < 0$). We did not try to optimize all model parameters simultaneously but instead used the separately optimized models described above. Note that more sophisticated methods exist for combining models using a Bayesian approach (Marzocchi *et al.*, 2012).

Averaging ETAS and kernel models allows us to combine the strengths of each model. In many cases the individual models provide very different results, with predicted rates varying by a factor of 100 (Fig. 6). The kernel and ETAS models are based on very different assumptions and thus have different weaknesses. Kernel models often perform better when seismicity rate is very active but varies smoothly, at scales much larger than typical interevent times and distances. This is generally the case during the first days following large mainshocks or during very active swarms. During these periods, ETAS model may be biased if the mainshock productivity or the aftershock decay with time are different from the prediction of the ETAS model with fixed parameters. On the other hand, kernel models may produce poor forecasts when the seismicity rate varies rapidly, at times smaller than the time window T, because it simply assumes that the rate during the next time window of duration T will be constant and equal to the present rate. Besides, model K^3 can account for

Table 3
Probability Gains G, Mixing Weight ω, and δAIC Values for Ensemble Models

			G			G	ω	δAIC	G	ω	δAIC	G	ω	δAIC
M_D	M_t	T (days)	ETAS	K^2	K^3	$\langle$ETAS, $K^2\rangle$			$\langle$ETAS, $K^3\rangle$			$\langle K^2, K^3\rangle$		
2	2	1	47.40	46.74	46.79	50.71	0.41	−21039	**51.89**	0.42	−28174	48.25	0.48	−32171
2	3	1	52.12	47.57	47.35	54.41	0.54	−1369	**54.49**	0.56	−1413	48.62	0.52	−4339
2	4	1	47.37	41.93	41.32	**48.59**	0.66	−73.2	48.24	0.72	−47.7	43.89	0.53	−443
2	5	1	23.36	22.14	22.08	**24.06**	0.60	0.8	23.95	0.66	4.3	22.50	0.55	−8.5
3	3	1	45.18	40.57	38.32	**47.35**	0.55	−1495	47.29	0.61	−1456	41.39	0.63	−4907
3	4	1	42.76	34.61	34.64	**43.55**	0.71	−50.1	43.47	0.74	−41.9	35.32	0.53	−729
3	5	1	22.38	18.72	20.68	22.60	0.79	6.9	**23.05**	0.63	2.8	20.70	0.10	−19.9
4	4	1	33.55	22.82	20.83	**34.14**	0.75	−47.1	33.91	0.84	−22.9	22.86	0.92	−1280
2	4	0.04	213.0	133.1	94.04	**216.2**	0.81	−38.4	214.26	0.93	−7.6	136.47	0.81	−1541
3	4	0.04	198.5	119.1	98.22	**205.5**	0.71	−103.	199.8	0.91	−9.3	125.2	0.82	−1678
2	4	10	16.69	13.87	14.64	16.73	0.86	0.9	**16.77**	0.81	−5.3	14.91	0.34	−431
3	4	10	15.97	12.84	13.67	16.06	0.87	−9.7	**16.16**	0.80	−27.8	13.73	0.24	−533
2	4	91	6.483	6.521	6.581	**6.749**	0.46	−98.3	6.745	0.43	−66.6	6.662	0.37	−66.6
3	4	91	6.823	6.301	6.387	6.823	1.00	10.0	**6.824**	0.96	11.5	6.404	0.29	−202

Probability gain G for different values of the minimum magnitude of target events M_t, of the minimum magnitude M_D of the learning catalog, and of the time interval T, computed for individual and ensemble models. Ensemble models (noted $\langle , \rangle$) are obtained by a weighted average of the rate of each individual model, with a weight ω for the first model and $1 - \omega$ for the second one. Bold values indicate the model with the largest gain for each set of parameters. A negative value of δAIC indicates that the ensemble model is better than each individual model.

deviations from the Gutenberg–Richter magnitude distribution or for temporal and spatial changes of b-value.

Discussion and Conclusion

We developed new methods for estimating the seismicity rate in space, time, and magnitude using adaptive kernels. Based on these estimates, we generated short-term, time-dependent earthquake forecasts that simply assume the seismicity rate over the forecast horizon equals the present rate, that is, these are persistence forecasts. Although this assumption seems simplistic, the models perform relatively well (albeit not generally better) in terms of likelihood compared with the ETAS model, even for forecast horizons of three months.

The kernel and ETAS models are based on very different assumptions. The ETAS model estimates the rate of background and triggered earthquakes, based on empirical laws describing the distribution of triggered earthquakes in time, space, and magnitude. In contrast, the kernel models are based on purely statistical methods and could be used to predict any (marked) space–time point process. Despite these fundamental differences between kernel models and ETAS model, the results (distribution in time, space, and magnitude, and probability gains) are rather similar. In terms of the likelihood score or the probability gain, ETAS model generally performs better than kernel models (see Table 2). However, using a different criterion, such as the mean predicted rate in nonempty bins, the kernel models perform better than ETAS in a majority of cases. Because of their simplicity and their lack of assumptions about seismicity, the kernel models can be used as a good null hypothesis against which to test other time-dependent models.

We optimized the ETAS parameters by maximizing the forecasts' likelihood under the assumption of a Poisson process in each bin, rather than by maximizing the likelihood function of the pointwise ETAS process. The parameters are thus effective. The ETAS model in definition (7) calculates the seismicity rate continuously as a function of all past events. The rate should thus be updated after each new earthquake to include all available information optimally; however, in our procedure, the learning catalog is only updated every T days. The effective ETAS parameters partially account for the contribution from earthquakes that occur during the forecast horizon. Also, our estimates of the productivity exponent α varies between 0.4 and 0.8, depending on learning and target catalog choices (M_D, M_t, T). These values are significantly smaller than the value $\alpha \approx 1$ obtained when directly estimating aftershock productivity using declustering methods (Helmstetter *et al.*, 2005; Hainzl and Marsan, 2008). Our implementation of the ETAS model can thus be considered as an intermediate model between a physical model (i.e., one that describes the rate of triggered earthquakes as a function of time and distance from the triggering earthquake) and a fully statistical model such as the kernel models (using kernels to measure the seismicity rate, without distinction between triggered earthquakes and other events).

Whereas the K^3 model does not assume a Gutenberg–Richter law (except for the small stationary background rate), it often approximates the magnitude distribution reasonably well. The probability gain G_M in magnitude (14) is, however, smaller than one for most cases shown in Table 2. This means that either the magnitude distribution obeys Gutenberg–Richter law at all time and space scales or that kernel smoothing is not a good approach for modeling the magnitude

distribution and its variation in time and space. Another approach would be to impose a Gutenberg–Richter law and to compute its b-value in each cell at each time step from the magnitudes of past nearby earthquakes.

Finally, we showed that constructing simple ensemble models from the kernel models and the ETAS model by simply averaging the predicted rate per bin often yields a higher predictive skill than each individual model can attain.

We submitted the ETAS and K^3 models to the CSEP testing center at the Southern California Earthquake Center (SCEC) in September 2012 for independent and prospective evaluation in the California testing area. There, our models compete with the models submitted by other researchers in the forecast group of daily forecasts of earthquakes with magnitudes $M_t \geq 3.95$. We submitted two versions of each model, one uses all prior earthquakes $M \geq 2$ as the learning catalog, whereas the other uses $M \geq 3$. For each learning catalog threshold, we also submitted a simple ensemble model that averages the forecasts of the ETAS and K^3 models.

Between the date of installation at CSEP and the date of the latest available test results (1 August 2013 at this time, 6 September 2013), 31 earthquakes $M \geq 3.95$ occurred within the testing region. Thus far, our models appear to perform well and as intended. Although it is too early to judge the practical significance of these results, all three models thus far perform better than any other installed one-day model. Within both the $M \geq 2$ and the $M \geq 3$ forecast group, K^3 achieved the highest gain, followed by the ensemble model and ETAS. According to the T- and W-tests, all three models currently achieve significantly larger gains than any other model, including the STEP model by Gerstenberger *et al.* (2005) and the critical-branching model by Kagan and Jackson (2010). Accumulating target earthquakes will enable more meaningful distinctions between the information content of the forecasts.

Data and Resources

We used the Advanced National Seismic System earthquake catalog made publicly available by the Northern California Earthquake Data Center at www.ncedc.org (last accessed March 2012) in the period from 1 January 1981 until 1 March 2012 with magnitude $M \geq 2$ and in the spatial region defined by the Regional Earthquake Likelihood Models collection region, defined in Table 2 by Schorlemmer and Gerstenberger (2007).

Acknowledgments

The authors thank the associate editor and two anonymous reviewers for carefully reading the manuscript and providing many constructive suggestions. This work has been partially funded by the EC Project REAKT FP7-282862. M. J. W. was partially supported by the Southern California Earthquake Center (SCEC), which is funded under National Science Foundation Cooperative Agreement EAR-1033462 and U.S. Geological Survey Cooperative Agreement G12AC20038. The SCEC contribution number for this paper is 1765.

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates: A square root law, *Ann. Stat.* **10,** 1217–1223.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Contr.* **19,** no. 6, 716–723.

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes, *Rev. Geophys.* **46,** RG4006, doi: 10.1029/2008RG000260.

Bird, P., and Y. Y. Kagan (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings, *Bull. Seismol. Soc. Am.* **94,** no. 6, 2380–2399.

Choi, E., and P. Hall (1999). Nonparametric approach to the analysis of space-time data on earthquake occurrences, *J. Comput. Graph. Stat.* **8,** 733–748.

Dodge, D. A., G. C. Beroza, and W. L. Ellsworth (1996). Detailed observations of California foreshock sequences: Implications for the earthquake initiation process, *J. Geophys. Res.* **101,** doi: 10.1029/96JB02269.

Eberhard, D. A. J., J. D. Zechar, and S. Wiemer (2012). A prospective earthquake forecast experiment in the western Pacific, *Geophys. J. Int.* **190,** no. 3, 1579–1592, doi: 10.1111/j.1365-246X.2012.05548.x.

Enescu, B., S. Hainzl, and Y. Ben-Zion (2009). Correlations of seismicity patterns in southern California with surface heat flow data, *Bull. Seismol. Soc. Am.* **99,** no. 6, 3114–3123, doi: 10.1785/0120080038.

Gerstenberger, M., S. Wiemer, L. M. Jones, and P. A. Reasenberg (2005). Real-time forecasts of tomorrow earthquakes in California, *Nature* **435,** 328–331.

Hainzl, S. (2013). Comment on "Self-similar earthquake triggering, Båths law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California" by P. M. Shearer, *J. Geophys. Res.* **118,** doi: 10.1002/jgrb.50132.

Hainzl, S., and D. Marsan (2008). Dependence of the Omori-Utsu law parameters on main shock magnitude: Observations and modeling, *J. Geophys. Res.* **113,** no. B10309, doi: 10.1029/2007JB005492.

Hauksson, E., J. Stock, K. Hutton, W. Yang, J. A. Vidal-Villegas, and H. Kanamori (2010). The 2010 M_w 7.2 El Mayor-Cucapah earthquake sequence, Baja California, Mexico and southernmost California, USA: Active seismotectonics along the Mexican Pacific margin, *Pure Appl. Geophys.* **168,** no. 8–9, 1255–1277.

Helmstetter, A., and M. J. Werner (2012). Adaptive spatio-temporal smoothing of seismicity for long-term earthquake forecasts in California, *Bull. Seismol. Soc. Am.* **102,** 2518–2529, doi: 10.1785/0120120062.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2005). Importance of small earthquakes for stress transfers and earthquake triggering, *J. Geophys. Res.* **110,** no. B05S08, doi: 10.1029/2004JB003286.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). Comparison of short-term and time-independent earthquake forecast models for southern California, *Bull. Seismol. Soc. Am.* **96,** doi: 10.1785/0120050067.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007). High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California, *Seismol. Res. Lett.* **78,** 78–86, doi: 10.1785/gssrl.78.1.78.

Kagan, Y. Y. (1997). Seismic moment-frequency relation for shallow earthquakes: Regional comparison, *J. Geophys. Res.* **102,** 2835–2852, doi: 10.1029/96JB03386.

Kagan, Y. Y., and D. D. Jackson (2010). Short- and long-term earthquake forecasts for California and Nevada, *Pure Appl. Geophys.* **167,** 685–692, doi: 10.1007/s00024-010-0073-5.

Kagan, Y. Y., and L. Knopoff (1987). Statistical short-term earthquake prediction, *Science* **236,** 1563–1567.

Kisslinger, C., and L. M. Jones (1991). Properties of aftershocks in southern California, *J. Geophys. Res.* **96,** 11,947–11,958.

Lippiello, E., C. Godano, and L. de Arcangelis (2012). The earthquake magnitude is influenced by previous seismicity, *Geophys. Res. Lett.* **39,** L05309, doi: 10.1029/2012GL051083.

Marsan, D., and O. Lengliné (2008). Extending earthquakes' reach through cascading, *Science* **319,** no. 5866, 1076–1079, doi: 10.1126/science.1148783.

Marzocchi, W., and A. M. Lombardi (2009). Real-time forecasting following a damaging earthquake, *Geophys. Res. Lett.* **36,** L21302, doi: 10.1029/2009GL040233.

Marzocchi, W., J. D. Zechar, and T. H. Jordan (2012). Bayesian forecast evaluation and ensemble earthquake forecasting, *Bull. Seismol. Soc. Am.* **102,** no. 6, 2574–2584, doi: 10.1785/0120110327.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika* **78,** no. 3, 521–530.

Ogata, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point processes, *J. Am. Stat. Assoc.* **83,** 9–27.

Rhoades, D. A., D. Schorlemmer, M. Gerstenberger, A. Christophersen, J. D. Zechar, and M. Imoto (2011). Efficient testing of earthquake forecasting models, *Acta Geophysica* **59,** no. 4, 728–747, doi: 10.2478/s11600-011-0013-5.

Schorlemmer, D., and M. Gerstenberger (2007). RELM testing center, *Seismol. Res. Lett.* **78,** no. 1, 30–36.

Schorlemmer, D., M. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007). Earthquake likelihood model testing. *Seismol. Res. Lett.* **78,** no. 1, 17–29.

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities, *Comm. Stat. A* **14,** 1123–1136.

Shearer, P. M. (2012). Self-similar earthquake triggering, Båths law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California, *J. Geophys. Res.* **117,** no. B06310, doi: 10.1029/2011JB008957.

Utsu, T., Y. Ogata, and S. Matsuura (1995). The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth* **43,** 1–33.

Vidale, J. E., K. L. Boyle, and P. M. Shearer (2006). Crustal earthquake bursts in California and Japan: Their patterns and relation to volcanoes, *Geophys. Res. Lett.* **33,** L20313, doi: 10.1029/2006GL027723.

Werner, M. J., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2011). High resolution long-term and short-term earthquake forecasts for California, *Bull. Seismol. Soc. Am.* **101,** no. 4, 1630–1648, doi: 10.1785/0120090340.

Werner, M. J., K. Ide, and D. Sornette (2011). Earthquake forecasting based on data assimilation: Sequential Monte Carlo methods for renewal point processes, *Nonlinear Process. Geophys.* **18,** 49–70, doi: 10.5194/npg-18-49-2011.

Wiemer, S., and D. Schorlemmer (2007). ALM: An asperity-based likelihood model for California, *Seismol. Res. Lett.* **78,** 134–140.

Wiemer, S., and M. Wyss (2002). Mapping spatial variability of the frequency-magnitude distribution of earthquakes, *Adv. Geophys.* **45,** 259–302.

Wiemer, S., S. R. McNutt, and M. Wyss (1998). Temporal and three-dimensional spatial analyses of the frequency-magnitude distribution near Long Valley caldera, California *Geophys. J. Int.* **134,** 409–421.

ISTerre
Université de Grenoble 1
CNRS, BP 53
F-38041 Grenoble, France
agnes.helmstetter@ujf-grenoble.fr
(A.H.)

Department of Geosciences
Princeton University
Princeton, New Jersey 08544
mwerner@princeton.edu
(M.J.W.)

Manuscript received 1 May 2013;
Published Online 18 February 2014